

PCTWORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification ⁶ : C12N 9/96, C07K 14/00, C12P 19/34	A1	(11) International Publication Number: WO 95/03401 (43) International Publication Date: 2 February 1995 (02.02.95)
(21) International Application Number: PCT/US94/08375 (22) International Filing Date: 25 July 1994 (25.07.94) (30) Priority Data: 08/097,173 23 July 1993 (23.07.93) US (71) Applicant: HYSEQ, INC. [US/US]; 670 Almanor Avenue, Sunnyvale, CA 94086 (US). (72) Inventor: GRUBER, Lewis, S.; Apartment 3911, 400 East Randolph, Chicago, IL 60601 (US). (74) Agent: MEYERS, Thomas, C.; Marshall, O'Toole, Gerstein, Murray & Borun, 6300 Sears Tower, 233 South Wacker Drive, Chicago, IL 60606 (US).		(81) Designated States: AU, BB, BG, BR, BY, CA, CN, CZ, FI, HU, JP, KP, KR, KZ, LK, LV, MG, MN, MW, NO, NZ, PL, RO, RU, SD, SK, UA, UZ, VN, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG). Published <i>With international search report.</i>
(54) Title: METHOD FOR SCREENING UNKNOWN ORGANISMS (57) Abstract A method for identifying an unknown organism by comparison of the nucleotide sequence of nucleic acid in a sample with known nucleotide sequences and a determination that an unknown nucleotide sequence is present in the sample.		

CGK00009421

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	GB	United Kingdom	MR	Mauritania
AU	Australia	GE	Georgia	MW	Malawi
BB	Barbados	GN	Guinea	NE	Niger
BE	Belgium	GR	Greece	NL	Netherlands
BF	Burkina Faso	HU	Hungary	NO	Norway
BG	Bulgaria	IE	Ireland	NZ	New Zealand
BJ	Benin	IT	Italy	PL	Poland
BR	Brazil	JP	Japan	PT	Portugal
BY	Belarus	KE	Kenya	RO	Romania
CA	Canada	KG	Kyrgyzstan	RU	Russian Federation
CF	Central African Republic	KP	Democratic People's Republic of Korea	SD	Sudan
CG	Congo	KR	Republic of Korea	SE	Sweden
CH	Switzerland	KZ	Kazakhstan	SI	Slovenia
CI	Côte d'Ivoire	LI	Liechtenstein	SK	Slovakia
CM	Cameroon	LK	Sri Lanka	SN	Senegal
CN	China	LU	Luxembourg	TD	Chad
CS	Czechoslovakia	LV	Latvia	TG	Togo
CZ	Czech Republic	MC	Monaco	TJ	Tajikistan
DE	Germany	MD	Republic of Moldova	TT	Trinidad and Tobago
DK	Denmark	MG	Madagascar	UA	Ukraine
ES	Spain	ML	Mali	US	United States of America
FI	Finland	MN	Mongolia	UZ	Uzbekistan
FR	France			VN	Viet Nam
GA	Gabon				

- 1 -

METHOD FOR SCREENING FOR UNKNOWN ORGANISMS

FIELD OF THE INVENTION

The present invention relates in general
5 to methods for screening for a nucleic acid of an
organism for which a nucleotide sequence is not
known, and in particular to methods employing
nucleotide sequencing for identification of
organisms.

10 BACKGROUND

Nucleotide sequencing provides sequence
information with various degrees of redundancy. The
information obtained from nucleotide sequencing may
be used as a source of primary sequence information
15 about the genomes of organisms, and once the
nucleotide sequence is known, may be used as a basis
for obtaining expression of sequenced genes and of
diagnosis of organisms containing the sequenced
genes. However, there are prospective advantages
20 for other uses for nucleotide sequencing which do
not require knowledge of the existence of the
organism to be sequenced.

SUMMARY OF THE INVENTION

The present invention provides a method
25 for screening a sample containing a nucleic acid for
the presence of an organism for which a nucleotide
sequence is not known including: sequencing all
nucleic acid in a sample; comparing the nucleotide
sequence obtained in sequencing step to nucleotide
30 sequences from known organisms; identifying a
continuous run of nucleotide sequence as not

- 2 -

corresponding to a known nucleotide sequence; and confirming the continuous run of nucleotide sequence as a nucleotide sequence of an organism for which the nucleotide sequence was not otherwise known.

5 A method according to the present invention may include a sequencing step including the step of sequencing the nucleic acid by hybridization with probes of known sequence.

10 Preferably a method according to the present invention includes a confirming step comprising the step of constructing an oligonucleotide probe having a continuous sequence of nucleotides or the complement thereto as found in the unknown sequence but not in known sequences;
15 exposing, under stringent hybridization conditions, the labeled oligonucleotide probe to a sample suspected of containing the oligonucleotide sequence; and identifying the presence of a previously hybridization complex between the labeled
20 oligonucleotide probe and nucleic acid in the sample. Stringent hybridization conditions are those understood in the art to result in hybridization of probes with perfectly matched, but not mismatched sequences of nucleotides.

25 A method according to the present invention may further comprise a second comparing step wherein a second continuous run of nucleotide sequence is compared with known nucleotide sequences.

30 A confirming step according to the present invention may comprise the step of: exposing the sample under stringent hybridization conditions to an oligonucleotide probe complementary to a portion of the unknown nucleotide sequence but not to a
35 known nucleotide sequence; and separating a fraction

- 3 -

containing a nucleic acid hybridizing to the labeled oligonucleotide from other fractions of the sample. A method according to the present invention may further include the step of microscopically
5 examining the fraction containing the labeled oligonucleotide probe, sequencing nucleic acid in the fraction containing the labeled oligonucleotide probe, and/or a second exposing step wherein the labeled oligonucleotide probe is exposed under
10 stringent hybridization conditions to a second sample.

A method according to the present invention may include a second exposing step comprising the step of obtaining a sample from a
15 second individual or a second sample from the same individual.

DETAILED DESCRIPTION

Nucleic acids and methods for isolating and cloning such nucleotide sequencing are well
20 known to those of skill in the art. See e.g., Ausubel et al., *Current Protocols in Molecular Biology*, Vol. 1-2, John Wiley & Sons Publs. (1989); and Sambrook et al., *Molecular Cloning A Laboratory Manual*, 2nd Ed., Vols. 1-3, Cold Spring Harbor Press
25 (1989), both of which are incorporated by reference herein.

Sequencing by hybridization ("SBH") is a well developed technology that may be practiced by a number of methods known to those skilled in the art.
30 Specifically, techniques related to sequencing by hybridization of the following documents is incorporated by reference herein: Drmanac et al., U.S. Patent No. 5,202,231 - Issued April 13, 1993; Drmanac et al., *Genomics*, 4, 114-128 (1989); Drmanac

- 4 -

et al., *Proceedings of the First Int'l. Conf. Electrophoresis Supercomputing Human Genome Cantor, DR & Lim HA eds, World Scientific Pub. Co., Singapore, 47-59 (1991)*; Drmanac et al., *Science*, 5 260, 1649-1652 (1993); Lehrach et al., *Genome Analysis: Genetic and Physical Mapping*, 1, 39-81 (1990), Cold Spring Harbor Laboratory Press; Drmanac et al., *Nucl. Acids Res.*, 4691 (1986); Stevanovic et al., *Gene*, 79, 139 (1989); Panuesku et al., *Mol. Biol. Evol.*, 1, 607 (1990); Drmanac et al., *DNA and Cell Biol.*, 9, 527 (1990); Nizetic et al., *Nucl. Acids Res.*, 19, 182 (1991); Drmanac et al., *J. Biomol. Struct. Dyn.*, 5, 1085 (1991); Hoheisel et al., *Mol. Gen.*, 4, 125-132 (1991); Strezoska et al., 10 *Proc. Nat'l. Acad. Sci. (USA)*, 88, 10089 (1991); Drmanac et al., *Nucl. Acids Res.*, 19, 5839 (1991); and Drmanac et al., *Int. J. Genome Res.*, 1, 59-79 (1992).

SBH technology may be applied to obtain 20 nucleotide sequence information for all or part of the genomes of known organisms. In this process, a number of oligonucleotide probes of a given length, which may be a 7-mer, are separately exposed under hybridization conditions with a sample to be 25 sequenced. Less than the total number of possible probes of a given length may be employed using various techniques, and exposure under hybridization conditions of probes of more than one length may be employed to improve the results. SBH may be 30 complimented by gel sequencing to obtain all of an unknown sample sequence.

According to the present invention, SBH may be applied to a sample of nucleic acid to determine whether it contains nucleic acid from at 35 one organism for which a nucleotide sequence is

- 5 -

unknown. Preferably, the sample may contain more than one genome. The nucleotide sequence obtained for a nucleic acid in the sample may be compared with nucleotide sequences for genomes of known organisms which may be eliminated from consideration. Continuous nucleotides sequence obtain from a sample, which sequence does not correspond to any known nucleotide sequence for a known organism, identifies the presence of a previously unknown organism.

The nucleotide sequence for the previously unknown organism that is obtained by SBH may then be used to make labeled oligonucleotide probes to diagnose the presence or absence of the organism and as an aid in identifying and isolating the previously unknown organism. Techniques such as filtration, centrifugation and chromatography may be applied to separate the organism from otherwise known organisms. Labeled oligonucleotide probes may be used as markers to identify the presence of the previously unknown organism in separatory fractions to obtain purified samples of organisms. Such purified samples of the organism may be sequenced in order to verify the original determination of the presence of a previously unknown organism and to verify the obtaining of a nucleotide sequence for the organism to whatever degree of completeness is desired.

Identification of previously unknown organisms by SBH may be employed in a diagnostic setting for determining organisms responsible for causing disease. Similarly, the method according to the present invention may be applied to identify new organisms in, for example, soil, air and water samples. Such a determination may be used to screen

- 6 -

for organisms having a desirable or undesirable effect observed from the soil, air or water sample (such as degradation of pollutants or nitrification). Similarly, organisms having an adverse or beneficial when found effect in food may be detected by using the method of the present invention. For example, where a phenotype is desired, a microorganism which has desirable properties may be identified by SBH even out of a mixture of unknown organisms by correlating presence of hybridization with a labeled probe constructed on the basis from SBH with the presence in a sample of the desired phenotype.

EXAMPLE

A blood sample from a subject exhibiting disease symptoms screened according to the present invention. Fractions of the blood sample suspected of containing a microorganism which may be responsible for the disease symptoms are preparatively treated to obtain a cDNA library useful for screening. Such preparation may include cloning of the DNA in vectors and amplification of the cloned nucleic acid by PCR.

After application of SBH procedures, sequence information is obtained. The sequence information is in the form of stretches of nucleotide sequence representing the overlapping runs of nucleotide sequence (a run being a continuous sequence formed by overlapping more than one probe sequence) of oligonucleotide probes which hybridize to cloned DNA from the sample. Nucleotide sequences known, e.g., from GENBANK (BBN Laboratories, Inc. 10 Moulton Street, Cambridge, MA) or another source of nucleotide sequence information

- 7 -

are excluded while the remaining sequences are further examined as follows.

In some instances, sequence from more than two clones may partially overlap, indicating the presence of a branch point. Such a branch point may indicate two similar stretches of nucleotide sequence in an organism in the sample, or may indicate a common portion of a sequence in two or more organisms in the sample. The sequences through each branch of the branch point are compared to known sequences. If the nucleotide sequence of a branch sequence does not correspond to a known nucleotide sequence for an organism, the determination of the nucleotide sequence of the branch is taken as an identification of an unknown organism.

Discontinuous runs of overlapping sequence which do not correspond to a nucleotide sequence from an organism for which a nucleotide sequence is known, may indicate a fragmentary sequence is present or may indicate that more than one organism is present. Such discontinuous runs of sequence are compared with nucleotide sequences from known organisms, and, to the extent that the sequence from the sample does not correspond to a known sequence, presence of at least one organism is identified.

The presence of an unknown organism is verified by synthesizing an oligonucleotide probe corresponding to a unique portion of a continuous run of nucleotide sequence identified as coming from an unknown organism or to the complement of the sequence. Such a probe is applied to the sample to confirm the presence of the determined sequence in the sample. Such a probe is applied to: another sample from the same individual from which the first

- 8 -

sample was derived; to a sample from a second individual who has diagnostic disease symptoms similar to those of the first individual; and to a sample from a third individual who does not have
5 diagnostic symptoms similar to those of the first individual. The presence of the nucleotide sequence but in samples from the same and the second individual but not the third individual identifies the sequence as being from a previously unknown
10 organism.

Oligonucleotide probes are made and used to identify a fraction of a sample from an individual identified as containing the nucleic acid above. The contents of the fraction hybridizing to
15 a labeled probe having the same or the complement of a nucleotide sequence of a previously unknown organism are examined using microscopic techniques to visually detect a previously unknown organism. Separatory techniques are applied to fractions
20 containing the previously unknown organism to track the presence of the previously unknown organism through fractions obtain from purification procedures known to those skilled in the art, which procedures separate the previously unknown organism
25 from known organisms in the sample.

Once separated from other organisms in the sample, sequencing by hybridization is applied to obtain a complete nucleotide sequence for the nucleic acid of the previous unknown organism.

30 The present invention has been described in terms of a particular embodiment. However, it is contemplated that modifications and improvements will occur to those skilled in the art upon consideration of the present specification and
35 claims. For example, although a preferred method

- 9 -

using SBH has been exemplified herein, gel sequencing or other nucleotide sequencing techniques may be employed solely or in combination with each other or SBH. Accordingly, it is intended that all
5 variations and modifications of the present invention be included within the scope of the claims.

- 10 -

CLAIMS

1. A method for screening a sample containing a nucleic acid for the presence of an organism for which a nucleotide sequence is not
5 known comprising the steps of:
 - sequencing all nucleic acid in a sample;
comparing the nucleotide sequence obtained
in said sequencing step to nucleotide sequences from
known organisms;
 - 10 identifying a continuous run of nucleotide
sequence as not corresponding to a known nucleotide
sequence; and
 - confirming the continuous run of
nucleotide sequence as a nucleotide sequence of an
15 organism for which the nucleotide sequence was not
otherwise known.
2. The method as recited in claim 1
wherein said sequencing step comprises the step of
sequencing the nucleic acid by hybridization with
20 probes of known sequence.

- 11 -

3. The method as recited in claim 1 wherein said confirming step comprises the step of constructing an oligonucleotide probe having a continuous sequence of nucleotides or the complement thereto as found in the unknown sequence but not in known sequences;

5 exposing, under stringent hybridization conditions, the labeled oligonucleotide probe to a sample suspected of containing the oligonucleotide sequence; and

10 identifying the presence of a hybridization complex between the labeled oligonucleotide probe and the previously unknown nucleic acid in the sample.

15 4. The method as recited in claim 1 further comprising a second comparing step wherein a second continuous run of nucleotide sequence is compared with known nucleotide sequences.

5. The method as recited in claim 1 wherein said confirming step comprises the steps of:

20 exposing the sample under stringent hybridization conditions to an oligonucleotide probe complementary to a portion of said unknown nucleotide sequence but not to a known nucleotide sequence; and

25 separating a fraction containing a nucleic acid hybridizing to the labeled oligonucleotide from other fractions of the sample.

6. The method as recited in claim 5 further comprising the step of microscopically examining the fraction containing the labeled oligonucleotide probe.

30

- 12 -

7. The method as recited in claim 5 further comprising the step of sequencing nucleic acid in the fraction containing the labeled oligonucleotide probe.

5 8. The method as recited in claim 5 further comprising a second exposing step wherein the labeled oligonucleotide probe is exposed under stringent hybridization conditions to a second sample.

10 9. The method as recited in claim 8 wherein said second exposing step comprises the step of obtaining a sample from a second individual.

15 10. The method as recited in claim 8 wherein said second exposing step comprises the step of obtaining a second sample from the same individual.

INTERNATIONAL SEARCH REPORT

Int. national application No.
PCT/US94/08375

A. CLASSIFICATION OF SUBJECT MATTER IPC(6) : C12N 9/96; C07K 14/00; C12P 19/34 US CL : 435/6, 188.5; 530/350 According to International Patent Classification (IPC) or to both national classification and IPC		
B. FIELDS SEARCHED Minimum documentation searched (classification system followed by classification symbols) U.S. : 435/6, 188.5; 530/350 Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched Electronic data base consulted during the international search (name of data base and, where practicable, search terms used) APS, MEDLINE, CA		
C. DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	KEETON, "BIOLOGICAL SCIENCE," published 1980 by W. W. NORTON & COMPANY, INC. (N.Y.), pages 893-897, especially pages 895 and 896.	1-10
<input type="checkbox"/> Further documents are listed in the continuation of Box C. <input type="checkbox"/> See patent family annex.		
* Special categories of cited documents: "A" document defining the general state of the art which is not considered to be of particular relevance "E" earlier document published on or after the international filing date "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) "O" document referring to an oral disclosure, use, exhibition or other means "P" document published prior to the international filing date but later than the priority date claimed "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art "&" document member of the same patent family		
Date of the actual completion of the international search 31 OCTOBER 1994		Date of mailing of the international search report NOV 09 1994
Name and mailing address of the ISA/US Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231 Facsimile No. (703) 305-3230		Authorized officer <i>SCOTT HOUTTEMAN</i> Telephone No. (703) 308-0196

Form PCT/ISA/210 (second sheet)(July 1992)*

CGK00009435